The Annals of Statistics
1985, Vol. 13, No. 2, 775–784

ASYMPTOTIC NUMBER OF ROOTS OF CAUCHY LOCATION LIKELIHOOD EQUATIONS[1]

BY JAMES A. REEDS

AT&T Bell Laboratories

The number of local maxima of the Cauchy location likelihood function which are not global maxima is asymptotically Poisson distributed with mean parameter $1/\pi$.

1. Result. Let $X_1, X_2, \cdots, X_n$ be iid random variables with density $1/\pi(1 + x^2)$ and let $R_n = R_n(X_1, \cdots, X_n)$ be the set of roots of the "Cauchy location likelihood equation":

$$R_n = \left\{\theta: \sum_{i=1}^n \frac{\partial}{\partial\theta} \log \frac{1}{\pi(1 + (X_i - \theta)^2)} = 0\right\}.$$

How many elements has R_n; how many roots are there? The standard maximum likelihood theory guarantees that one of the elements of R_n—the maximum likelihood estimate—is close to 0, and that all other elements are bounded away in probability as $n \to \infty$ from 0. But it does not tell us how many elements there are.

Let $r_n = \text{card}(R_n)$ be the number of roots. With probability one the likelihood equation has only simple roots, which are alternately local maxima and minima of the likelihood function. Hence r_n is odd, there are $\frac{1}{2}(r_n + 1)$ local maxima and $\frac{1}{2}(r_n - 1)$ local minima. Of the local maxima, one is the global maximum: the maximum likelihood estimate. Let us call the other local maxima "false maxima". These $\frac{1}{2}(r_n - 1)$ false maxima are an embarrassment for the maximum likelihood method of estimation; the following theorem shows that their number is really quite small.

THEOREM 1. *If* $X_1, X_2, \cdots$ *are iid with density* $1/\pi(1 + x^2)$, *then for each* k,

$$\lim_{n\to\infty} P(\tfrac{1}{2}(r_n - 1) = k) = e^{-(1/\pi)}/\pi^k k!.$$

That is, the number of false maxima of the Cauchy likelihood function is asymptotically Poisson with parameter $1/\pi \approx .31831$. This agrees with computer experiment results obtained by Barnett (1966), part of whose data are given in Table I.

2. Explanation. This result rests on the fact that in the limit all the local maxima are in one-to-one correspondence with those data points which are

Received January 1979; revised January 1984.

[1] This research was partially supported by National Science Foundation Grant No. MCS75-10376-A02.

AMS 1980 *subject classifications.* Primary 62-00, 62E20; secondary 62F12.

Key words and phrases. Cauchy likelihood equations, local maxima, fluctuation inequality, Poisson distribution.

TABLE I
V. D. Barnett's Monte Carlo Results

n	Monte Carlo sample size	Empirical relative frequency of number of local maxima $(r_n - 1)/2$ 1	2	3	4	5
3	18000	.646	.262	.092	...	...
5	3000	.652	.268	.069	.011	.001
7	3000	.670	.261	.058	.009	.001
9	2250	.673	.269	.052	.005	...
11	1000	.706	.245	.036	.012	.001
13	931	.698	.255	.039	.009	...
15	1000	.696	.262	.039	.002	.001
19	784	.707	.245	.043	.005	...
∞	Exact limiting distribution	.727377	.231531	.036849	.003910	.000311

greater than $2n$ in absolute value. In particular, the proof of the theorem is broken into the following steps:

1. If θ is a local maximum there is some i for which $|X_i - \theta| \leq 1$. We may think, then, of each observation X_i as having a nearby local maximum *in potentiam*. Depending on the configuration of the other observations, this potential local maximum may or may not become an *actual* local maximum. If there are several observations X_j close to X_i, then the potential local maximum at X_i will not be expressed; if the observation X_i is isolated from the other observations, its potential maximum is likely to manifest itself as an actual maximum. We will see that the false maxima will typically occur in the sparse outlying fringes of the observed scatter of the data X_i and not in the crowded central portion.

2. The classical Wald theory of the consistency of the maximum likelihood estimate rules out the occurrence (in probability) of any false maxima in the range $[-e, e]$. This is explained in Perlman (1983).

3. The central limit theorem, coupled with a fluctuation inequality for the sample score function, makes the number of false maxima in the ranges $[-n(\log n)^{-A}, -e]$ and $[e, n(\log n)^{-A}]$ (where A is a certain fixed positive constant) tend to zero (in probability).

4. Chebyshev's inequality, applied to the sample score function, shows that the number of false maxima in the ranges $[n^{3/4}, 2n - \sqrt{n}]$ and $[-2n + \sqrt{n}, -n^{3/4}]$ tends to zero in probability.

5. Rouché's theorem and Chebyshev's inequality show that with probability tending to unity each observation greater than $2n - \sqrt{n}$ in absolute value has associated with it a unique false maximum.

6. But the number of such observations is exactly a binomial random variable

with parameters n and

$$p_n = \frac{2}{\pi} \int_{2n-\sqrt{n}}^{\infty} \frac{1}{1+x^2}\, dx \approx \frac{1}{\pi n},$$

and approximately Poisson with parameter $\lambda = 1/\pi$.

3. Proofs. For ease of presentation, each of the steps of the outline in the preceding section is given as a separate proposition, each with its own little proof. First, however, general notations. Let $\psi(s) = 2s/(1+s^2)$ and let $\bar{\psi}(\theta) = (1/n) \sum_{i=1}^{n} \psi(X_i - \theta)$. Let $a \lesssim b$ be shorthand for $a = O(b)$. Let $a \approx b$ be shorthand for "$a \lesssim b$ and $b \lesssim a$". The set of roots $\bar{\psi}$ is the set R_n; the downcrossings of zero by $\bar{\psi}$ are the local maxima of the likelihood function. Elementary but tedious calculations yield

$$(1) \quad E\bar{\psi}(s) = -\frac{2s}{s^2+4}, \quad n \operatorname{Cov}(\bar{\psi}(s), \bar{\psi}(t)) = -\frac{8(|s-t|^2 - st - 4)}{(s^2+4)(t^2+4)(|s-t|^2+4)}$$

and

$$n \operatorname{Var}(\bar{\psi}(s)) = 2/(s^2+4).$$

Let $\psi_i^*(\theta) = \sum_{i \neq j} \psi(X_j - \theta)$ so that for each i,

$$\bar{\psi}(\theta) = (1/n)\psi_i^*(\theta) + (1/n)\psi(X_i - \theta).$$

Then

$$E\psi_i^*(\theta) = 2(n-1)\theta/(\theta^2+4) \approx -2n/\theta$$

if n and θ are big, and

$$\operatorname{Var} \psi_i^*(\theta) = 2(n-1)/(\theta^2+4) \approx 2n/\theta^2$$

if n and θ are big.

PROPOSITION 1. *If θ is a local maximum there is some i for which $|X_i - \theta| \le 1$.*

PROOF. If θ is a local maximum of the likelihood function, it is a downcrossing of zero by $\bar{\psi}$, and hence $(\partial/\partial\theta)\bar{\psi}(\theta) \le 0$, which can happen only if, for some i, $(\partial/\partial\theta)\psi(X_i - \theta) = -\psi'(X_i - \theta) \le 0$. But $\psi'(s) = 2(1-s^2)/(1+s^2)^2$ which is ≥ 0 only if $|s| \le 1$, so $|X_i - \theta| \le 1$.

PROPOSITION 2. (Perlman, 1983). *For each finite, positive K,*

$$\operatorname{card}([-K, K] \cap R_n) \to 1 \quad a.s.$$

PROOF. Banach space strong law of large numbers.

PROPOSITION 3. *There is a positive constant A such that the number of roots of $\bar{\psi}$ in the ranges $[e, n(\log n)^{-A}]$ and $[-n(\log n)^{-A}, -e]$ tends to zero in probability.*

PROOF. We will exhibit a function $\lambda(\theta)$ such that $\lambda(\theta)\psi(X_1 - \theta)$ obeys the central limit theorem in the Banach space $C[e, \infty]$ i.e., so that $\sqrt{n}(\lambda\bar{\psi} - E\lambda\bar{\psi})$ converges in distribution to a continuous Gaussian process on $[e, \infty]$. A similar argument applies on $[-\infty, -e]$.

Then

$$\lim_{n\to\infty} P(\sup_{e\le\theta\le\infty} |\lambda(\theta)\bar{\psi}(\theta) - E\lambda(\theta)\bar{\psi}(\theta)| \le \log n/\sqrt{n}) = 1.$$

Hence, with probability tending to unity as $n \to \infty$, the following argument applies: For any root θ^* of $\bar{\psi}$, we have

$$\lambda(\theta^*)|E\bar{\psi}(\theta^*)| = \lambda(\theta^*)2|\theta^*|/(4 + |\theta^*|^2) \le \log n/\sqrt{n},$$

and hence $|\theta^*| \ge k(n)$ where

$$k(t) = \inf\{x > 0 : 2\lambda(x)x/(4 + x^2) \le \log t/\sqrt{t}\}$$

is a function whose asymptotics may easily be derived in terms of λ. It turns out that $\lambda(\theta) = \theta^{1/2}(\log \theta)^{-3}$ will do, and then $k(t)$ may be underestimated by $t(\log t)^{-10}$. We may thus take $A = 10$ in the statement of the proposition.

To see that the central limit theorem applies, we verify a fluctuation inequality. According to Hahn (1977), it suffices to check that

$$(2) \qquad \gamma(s, t) = \operatorname{Var}(\lambda(s)\psi(X_1 - s) - \lambda(t)\psi(X_1 - t)) \lesssim f(|\eta(s) - \eta(t)|)$$

where η is a homeomorphism of $[e, \infty]$ onto $[0, 1]$ and where $f(y)$ is a nonnegative, nondecreasing function which satisfies $\int_0^1 y^{-3/2} f^{1/2}(y)\, dy < \infty$. In particular, let $\eta(s) = 1/\log s$ and let $f(y) = y/\phi(y)$, where $\phi(y)$ is 1 if $y > 1/e$ and $\phi(y) = |\log y|^3$ if $y \le 1/e$. Using formula (1) we see that

$$\gamma(s, t) = \frac{2|s - t|^2}{|s - t|^2 + 4}\left\{\frac{\lambda^2(s)}{s^2 + 4} + \frac{\lambda^2(t)}{t^2 + 4} + \frac{8\lambda(s)\lambda(t)}{(s^2 + 4)(t^2 + 4)}\right\}$$

$$+ \frac{2}{(s^2 + 4)(t^2 + 4)(|s - t|^2 + 4)}\{16(\lambda(s) - \lambda(t))^2 + 4(t\lambda(s) - s\lambda(t))^2\}.$$

A few terms in this big expression dominate the others; to see which they are we use the following properties of λ:

(i) $\lambda'(s) \lesssim \lambda(s)/s$. (ii) $\frac{d}{ds}(\lambda(s)/s) \lesssim \lambda(s)/s^2$.

(iii) $\lambda'(s)$ is decreasing. (iv) $\frac{d}{ds}(\lambda(s)/s)$ is decreasing.

Thus

$$|\lambda(s) - \lambda(t)|^2 = |s - t|^2|\lambda'(u)|^2 \le |s - t|^2\{|\lambda'(s)|^2 + |\lambda'(t)|^2\}$$
$$\lesssim |s - t|^2\{|\lambda(s)/s|^2 + |\lambda(t)/t|^2\}$$

for some u between s and t.

Similarly, for some u between s and t,

$$|t\lambda(s) - s\lambda(t)|^2 = s^2t^2\left|\frac{\lambda(s)}{s} - \frac{\lambda(t)}{t}\right|^2 = s^2t^2|s-t|^2\left|\frac{d}{du}\frac{\lambda(u)}{u}\right|^2$$

$$\lesssim s^2t^2|s-t|^2\left\{\frac{\lambda(s)^2}{s^4} + \frac{\lambda(t)^2}{t^4}\right\}.$$

As a result,

$$\gamma(s, t) \lesssim \frac{2|s-t|^2}{|s-t|^2+4}\left\{\frac{\lambda^2(s)}{s^2+4} + \frac{\lambda^2(t)}{t^2+4} + \frac{8\lambda(s)\lambda(t)}{(s^2+4)(t^2+4)}\right\}$$

and even

$$\gamma(s, t) \lesssim \frac{2|s-t|^2}{|s-t|^2+4}\left\{\frac{\lambda^2(s)}{s^2+4} + \frac{\lambda^2(t)}{t^2+4}\right\}.$$

Since $1/\log t$ decreases, since $\lambda(t)/t$ decreases, and since

$$\frac{|s-t|^2}{|s-t|^2+4} \approx \frac{\min(1, |s-t|)^2}{1+\min(1, |s-t|)^2}$$

it suffices to check that

$$\frac{\theta^2}{1+\theta^2}\frac{\lambda^2(s)}{s^2+4} \lesssim f\left(\frac{1}{\log(s)} - \frac{1}{\log(s+\theta)}\right)$$

for all $s \geq e$ and all θ in $[0, 1]$.

Let $\alpha(s) = \frac{1}{100}(1/(s\log(s)^2))$. Then it is easy to verify that

$$\theta\alpha(s) \leq \frac{1}{\log(s)} - \frac{1}{\log(s+\theta)}$$

for all $s \geq e$ and all θ in $[0, 1]$. Since f is nondecreasing it suffices to check that

$$\frac{\theta^2}{1+\theta^2}\frac{\lambda^2(s)}{s^2} \lesssim f(\theta\alpha(s)) = \frac{\theta\alpha(s)}{\phi(\theta\alpha(s))}$$

viz., that

$$\phi(\theta\alpha(s))(\theta/(1+\theta^2))\lambda^2(s)/s^2 \lesssim \alpha(s).$$

From the definition of $\alpha(s)$ and of $\lambda(s)$ it suffices to check that

$$\phi(\theta\alpha(s))\theta/(1+\theta^2) \lesssim (\log s)^4.$$

To check this last inequality note that $\phi(x) = \max(1, |\log x|^3) \lesssim 1 + |\log x|^3$ and so $\phi(\theta\alpha(s)) \lesssim 1 + |\log\theta|^3 + |\log\alpha(s)|^3$ and so it suffices to check that

$$(\theta/(1+\theta^2))(1 + |\log\theta|^3 + |\log\alpha(s)|^3) \lesssim (\log s)^4.$$

Trivially

$$(\theta/(1+\theta^2))(1+|\log\theta|^3) \lesssim 1 \lesssim (\log s)^4$$

so it finally suffices to check that

$$|\log\alpha(s)|^3 \lesssim (\log s)^4$$

which is clearly true by the definition of $\alpha(s)$.

To prove Proposition 4 we work with the functions ψ_i^*. Recall that

$$E\psi_i^*(\theta) = -\frac{2(n-1)\theta}{\theta^2+4} \approx -\frac{2n}{\theta}$$

if n and θ are big, and

$$\operatorname{Var}\psi_i^*(\theta) = \frac{2(n-1)}{\theta^2+4} \approx \frac{2n}{\theta^2}$$

if n and θ are big. If $\sqrt{n} < s < 2n - \sqrt{n}$ then

$$E\psi_i^*(s) < -1,$$

and if $s > 2n + \sqrt{n}$ then

$$-1 < E\psi_i^*(s) < 0.$$

In either case

$$(1 + E\psi_i^*(s))^2 \approx (2n-s)^2/s^2.$$

Hence, by Chebyshev's inequality, if $\sqrt{n} < s < 2n - \sqrt{n}$,

$$P(\psi_i^*(s) \geq -1) \lesssim 2n/(2n-s)^2,$$

and if $s > 2n + \sqrt{n}$,

$$P(\psi_i^*(s) \leq -1) \lesssim 2n/(2n-s)^2,$$

and

$$P(\psi_i^*(s) \geq 1) \leq P(\psi_i^*(s) \geq 0) \lesssim \frac{2n}{s^2}\frac{s^2}{4n^2} \approx \frac{1}{2n}.$$

Similar formulae hold for negative values of s.

Proposition 4. *Let A_i be the event that $n^{3/4} \leq |X_i| \leq 2n - \sqrt{n}$ and that there is a false maximum within distance* 1 *of X_i. Then $P(A_1 \cup A_2 \cup \cdots A_n) \to 0$ as $n \to \infty$.*

Proof. Suffices to show that $nP(A_1) \to 0$ as $n \to \infty$. Let $A(x)$ be the event that there is no false maximum within distance 1 of x. Let $B(x)$ be the event that $|X_j - x| > 1$ for all $j > 1$. Let $C(x)$ be the event that $\psi_1^*(x+1) < -1$ if $x > 0$ and that $\psi_1^*(x-1) > 1$ if $x < 0$. For all x the intersection of $B(x)$ and $C(x)$ is contained in $A(x)$ and so

$$1 - P(A(x)) \le 1 - P(B(x)) + 1 - P(C(x)).$$

Thus

$$P(A_1) = \frac{1}{\pi}\int_S \frac{dx}{1+x^2} P(\text{not } A(X_1) \mid X_1 = x) = \frac{1}{\pi}\int_S \frac{1 - P(A(x))}{1+x^2}\,dx$$
$$\lesssim \int_S \frac{1-P(B(x))}{1+x^2}\,dx + \int_S \frac{1-P(C(x))}{1+x^2}\,dx$$

where $S = \{x\colon n^{3/4} \le |x| \le 2n - \sqrt{n}\}$.

But

$$1 - P(B(x)) = 1 - \left(1 - \frac{1}{\pi}\int_{x-1}^{x+1} \frac{dt}{1+t^2}\right)^{n-1}$$
$$\approx 1 - \left(1 - \frac{2}{\pi x^2}\right)^{n-1} \approx \frac{2(n-1)}{\pi x^2} \lesssim \frac{2n}{x^2}$$

since $n \ll x^2$.

Similarly, if $x > 0$,

$$1 - P(C(x)) = P(\psi_1^*(x+1) \ge -1) \lesssim 2n/(2n-x)^2$$

and if $x < 0$

$$1 - P(C(x)) \lesssim 2n/(2n+x)^2$$

so

$$P(A_1) \lesssim \int_S \frac{dx}{1+x^2}\left(\frac{2n}{x^2} + \frac{2n}{(2n-|x|)^2}\right) = o\left(\frac{1}{n}\right).$$

PROPOSITION 5. *The expected number of observations X_i such that $2n - \sqrt{n} \le |X_i|$, which do not have precisely one false maximum within distance* 1 *of X_i, tends to* 0 *as $n \to \infty$.*

PROOF. We use Rouché's theorem to count roots in the complex plane. Let γ be the contour in the complex plane formed by the two circular arcs $\{\sqrt{2}e^{i\theta} - i\colon \pi/4 \le \theta \le 3\pi/4\}$ and $\{\sqrt{2}e^{i\theta} + i\colon 5\pi/4 \le \theta \le 7\pi/4\}$. γ encloses the origin, is symmetric with respect to complex conjugation, and $|\psi(s)| = 1$ for all $s \in \gamma$. Let $\gamma + x$ denote the translate of γ by x so that $\gamma + x = \{s \in \mathbb{C}\colon s - x \in \gamma\}$. Let $R(\delta, f)$ denote the number of roots of f enclosed by the contour δ. Since the rational function $\bar{\psi}(s)$ has real coefficients its complex roots occur in complex conjugate pairs. Hence if $R(\gamma + x, \bar{\psi}) = 1$ the root of $\bar{\psi}$ enclosed by $\gamma + x$ is real.

Say the observation X_i is *bad* if $X_i > 2n - \sqrt{n}$ and if there is not exactly one root within distance 1 of X_i. Let B_n be the number of such *bad* X_i; we show that

$EB_n \to 0$. Clearly

$$EB_n \le \sum_{i=1}^n \frac{1}{\pi} \int_{2n-\sqrt{n}}^{\infty} \frac{1}{1+x^2} P(R(\gamma + X_i, \bar{\psi}) \ne 1 \mid X_i = x)\, dx$$

$$\lesssim n \int_{2n-\sqrt{n}}^{\infty} \frac{1}{x^2} P(R(\gamma + X_1, \bar{\psi}) \ne 1 \mid X_1 = x)\, dx.$$

We use Rouché's theorem to estimate the probability that $R(\gamma + X_1, \bar{\psi}) \ne 1$. Compare the function $\psi(X_1 - s)$ with $n\bar{\psi}(s)$ on the contour $\gamma + X_1$. If for all $s \in \gamma + X_1$ we have $|\psi(X_1 - s)| > |n\bar{\psi}(s) - \psi(X_1 - s)| = |\sum_{i=2}^n \psi(X_i - s)|$ then

$$R(\gamma + X_1, \bar{\psi}) = R(\gamma + X_1, \psi(X_1 - s)) = R(\gamma, \psi(-s)) = 1.$$

But $|\psi(X_1 - s)| = 1$ for all $s \in \gamma + X_1$ so

$$\begin{aligned} P(R(\gamma + X_1, \bar{\psi}) &\ne 1 \mid X_1 = x) \\ &\le P(\sup_{s\in\gamma+X_1} |n\bar{\psi}(s) - \psi(X_1 - s)| \ge 1 \mid X_1 = x) \\ &= P(\sup_{s\in\gamma+x} |\textstyle\sum_{j=2}^n \psi(X_j - s)| \ge 1) \\ &\le P(\textstyle\sum_{j=2}^n \sup_{s\in\gamma} |\psi(X_j - s + x)| \ge 1). \end{aligned}$$

Let $Y_n(x) = \sum_{j=2}^n \sup_{s\in\gamma} |\psi(X_j - s + x)|$. We have shown that

$$EB_n \lesssim \int_{2n-\sqrt{n}}^{\infty} \frac{n}{x^2} P(Y_n(x) \ge 1)\, dx = \int_{2-1/\sqrt{n}}^{\infty} \frac{1}{t^2} P(Y_n(nt) \ge 1)\, dt.$$

We show in Lemma 3 that for each $t \ne 0$ the random variable $Y_n(nt)$ converges in probability to $2/|t|$, and hence for values of $t > 2$, $\lim_{n\to\infty} P(Y_n(nt) \ge 1) = 0$. Applying the dominated convergence theorem, we see that $\lim_{n\to\infty} EB_n = 0$.

This shows that with probability tending to 1 all observations greater than $2n - \sqrt{n}$ in absolute value are within distance 1 of exactly one root. The same calculation also shows that these roots are downcrossings, as follows. We just saw that with probability tending to 1 for all X_i with $|X_i| > 2n - \sqrt{n}$ we have

$$|n\bar{\psi}(s) - \psi(X_i - s)| < |\psi(X_i - s)|$$

for all $s \in \gamma + X_i$, including $s = X_i \pm 1$. Thus $\bar{\psi}(s)$ has the same sign as $\psi(X_i - s)$ for $s = X_i \pm 1$. Since $\psi(1) = 1$ and $\psi(-1) = -1$ we know $\bar{\psi}(X_i - 1) > 0 > \bar{\psi}(X_i + 1)$ and hence the root near is X_i is a downcrossing.

LEMMA 1. *Let X be a Cauchy random variable. Then*

$$\lim_{s\to\infty} E|X|/(1 + |X - s|) = 0.$$

PROOF. Easy.

LEMMA 2. *Let $f(x) = g(x)/(1 + |x|)$ be bounded. Suppose $\lim_{|x|\to\infty} g(x) = \alpha$. Let X be Cauchy. Then $\lim_{n\to\infty} Enf(X - nt) = \alpha/|t|$.*

PROOF. $|x|f(x)$ is bounded so the dominated convergence theorem shows $E|X - nt|f(X - nt) \to \alpha$. By Lemma 1,

$$E|X|f(X - nt) = E\frac{|X|}{1 + |X - nt|}g(X - nt) \lesssim E\frac{|X|}{1 + |X - nt|} \to 0$$

so $E|nt|f(X - nt) \to \alpha$.

LEMMA 3. *If $t \neq 0$ then $Y_n(t) \to 2/|t|$ in probability.*

PROOF. Let $f(x) = \sup_{s\in\gamma}|2(s - x)/(1 + (s - x)^2)|$. Then $g(x) = (1 + |x|)f(x)$ is bounded and $\lim_{|x|\to\infty}g(x) = 2$. But $Y_n(x) = \sum_{j=2}^n f(X_j - x)$ so $EY_n(nt) = (n - 1)Ef(X_1 - nt) \to 2/|t|$ by Lemma 2. Further,

$$\text{Var } Y_n(nt) = (n - 1)\text{Var } f(X_1 - nt) \leq (n - 1)Ef^2(X_1 - nt) \to 0$$

by another application of Lemma 2.

PROPOSITION 6. *The distribution of the number of observations exceeding $2n - \sqrt{n}$ in absolute value converges to the Poisson distribution with parameter $1/\pi$.*

PROOF. The number of such observations has a binomial distribution, with expectation n times the symmetric tail

$$\frac{1}{\pi}\left(\int_{-\infty}^{-2n+\sqrt{n}} + \int_{2n-\sqrt{n}}^{\infty}\right)\frac{1}{1 + x^2}\,dx.$$

Each of the two integrals is clearly asymptotically proportional to $1/2n$; thus the expectation of the number of such observations tends to $1/\pi$.

4. Poisson process limit. The proof of Theorem 1 actually yields more detailed information than claimed in Theorem 1. Let ν be the measure with Lebesgue density function $p(x) = (1/\pi)(1/x^2)$ if $|x| > 2$, and $p(x) = 0$ otherwise. Let $N_n(t)$ be the number of false maxima less than nt.

THEOREM 2. *As $n \to \infty$, the process $N_n(t)$ converges in distribution to the Poisson process on the real line with expectation measure ν.*

5. Generalizations. Results similar to the present ones can be derived in many other cases. One can both replace the Cauchy distribution of the data by some other iid model and replace the maximum likelihood method of estimation by another M-estimate method. In many cases simpler arguments than the ones used here are sufficient to count the number of false maxima. Typically the law of large numbers methods alone are sufficient. Occasionally the central limit theorem method of Proposition 3 is also needed. It is felt that the present Cauchy maximum likelihood case is harder than most of the other cases of interest to statisticians.

6. Acknowledgments. I am grateful to Bill Lepowsky for detecting an error in an earlier version of Proposition 5, to Larry Shepp for suggesting Lemma 3 and Theorem 2, to Richard Fateman for his computer derivation of formula (1) on the Project MAC MACSYMA symbol manipulation computer at MIT, and to Erich Lehmann and Michael Perlman for their encouragement in writing this paper.

REFERENCES

[1] BARNETT, V. D. (1966). Evaluation of the maximum likelihood estimator where the likelihood equation has multiple roots. *Biometrika* **53** 151–165.

[2] HAHN, M. (1977). Conditions for sample-continuity and the central limit theorem. *Ann. Probab.* **5** 351–360.

[3] PERLMAN, M. D. (1983). The limiting behavior of multiple roots of the likelihood equation. In *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on his Sixtieth Birthday* 339–370. (M. Rizvi, J. Rustagi, D. Siegmund, eds.) Academic, New York.

MATHEMATICS AND STATISTICS RESEARCH CENTER
AT&T BELL LABORATORIES
MURRAY HILL, NEW JERSEY 07974